Biofutures

Study Leaders:
D. Nelson
T. Hwa

Contributors Include:
H. Abarbanel
S. Block
F. Dyson
G. Joyce
J. Kimble
S. Koonin
N. Lewis
H. Levine
M. Prentiss
H. Woodin

June 2001

JSR-00-130

JASON
The MITRE Corporation
1820 Dolley Madison Boulevard
McLean, Virginia 22102-3481
(703) 883-6997

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 2001	

4. TITLE AND SUBTITLE

Biofutures

5. FUNDING NUMBERS

13-988534-A4

6. AUTHOR(S)

D. Nelson, T. Hwa et al.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The MITRE Corporation
JASON Program Office
1820 Dolley Madison Blvd
McLean, Virginia 22102

8. PERFORMING ORGANIZATION REPORT NUMBER

JSR-00-130

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Advanced Research Projects Agency
3701 N. Fairfax Drive
Arlington, Virginia 22203-1714

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

JSR-00-130

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release - Distribution Unlimited

12b. DISTRIBUTION CODE

Distribution Statement A

13. ABSTRACT *(Maximum 200 words)*

The goal of the 2000 JASON summer study on ``Biofutures" was to explore prospects for computer modeling of cellular biochemical networks and to ask more generally about the role of modeling in biology. Indeed, it is far from clear that reductionist models from the physical sciences are the right paradigm -- methodologies taken from operations research or electrical engineering may be more relevant [2]. Comments on modeling from biologists range from complaints about the lack of biological realism to categorical statements that groups who do not buy into these developments are going to be left behind. A majority of biologists are quite skeptical about the utility of models, notwithstanding the success of, for example, the Hodgkin-Huxley model of electrical impulses in nerve cells.

14. SUBJECT TERMS	15. NUMBER OF PAGES
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

Contents

1 OVERVIEW

The genomes of over 36 organisms (many of them simple unicellular prokaryotes) are now known. Among other applications, this information could be used to construct computer ("in silico") models of organisms. Indeed, efforts are already underway to model Mycoplasma genitalium (the smallest known genome of a free-living organism, with approximately 500 genes) by a set of coupled nonlinear rate equations for the concentrations of various structural proteins, enzymes, genes and messenger RNA's [1]. A successful effort of this kind would allow predictions of the response of simple organisms to drugs, changes in the environment or gene knockouts. Such models could also be an aid to rational drug design and could perhaps even guide biological research. Detailed models of specific biochemical networks (as opposed to whole cells) could aid in the design of cells which sense harmful factors in the environment or allow switching from one type of behavior to another.

The goal of the 2000 JASON summer study on "Biofutures" was to explore prospects for computer modeling of cellular biochemical networks and to ask more generally about the role of modeling in biology. Indeed, it is far from clear that reductionist models from the physical sciences are the right paradigm – methodologies taken from operations research or electrical engineering may be more relevant [2]. Comments on modeling from biologists range from complaints about the lack of biological realism to categorical statements that groups who do not buy into these developments are going to be left behind. A majority of biologists are quite skeptical about the utility of models, notwithstanding the success of, for example, the Hodgkin-Huxley model of electrical impulses in nerve cells.

There are real questions about what constitutes a good model in biology, and serious worries about "garbage in, garbage out" when one tries to represent a living cell as a system of hundreds of coupled nonlinear differential equations with thousands of poorly known rate constants. On the other hand, after efforts to incorporate important experimental facts, good modelers might be able put their finger on crucial parameters or nodes in biochemical pathways and identify less important ones. Although useful whole cell models are unlikely in the near future (even for prokaryotes), modeling isolated biological "modules" [2] (such as the lysis/lysogenesis switch in lambda-phage or chemotaxis in E-coli) seems promising. Understanding such modules, especially if they are conserved across species, could lead to a biological "tool kit" for developing specialized designer cells in silico before realizing them in the laboratory. Such cells could ultimately serve as more efficient "canary sensors", specialized to particular diseases or toxins. Someday, drug companies will surely want to design and test their products on cells in silico before going on to real animal models, just as airplane designers now simulate a new aircraft wing on a computer before subjecting it to wind tunnel experiments. The complexity of general circulation models of the earth's climate rivals the problems faced by modelers of entire cells, especially eukaryotes. Cellular modeling, however, has the advantage that it can be informed by experiments that probe the response of numerous genetically identical cells to different environmental stimulii.

Future generations of models will have to find a way to properly incorporate *spatial* variations in reactant concentrations, which are known to be important for certain processes, even in prokaryotic cells. For biochemical pathways that involve small numbers of molecules, stochastic simulations instead of deterministic ones can also be important. As a paradigm for

modeling that includes spatial effects, we mention the SPICE program (Simulated Program for Integrated Circuit Evaluation), developed for integrated circuit design in the 1970's. That program has played a key role in designing new generations of silicon chips. The hope is that independent SPICE-like programs will not be needed for each biological organism or specialized eukaryotic cell.

We will denote complex attempts to get all the details right, similar to the SPICE program, as"Type A" modeling. In solid state physics, Type A modeling is represented in solid state physics by ab initio electron band structure calculations. "Type B" modeling, where one tries to simplify and uncover general principles, is represented by tight binding models which illustrate the idea of energy bands and reveal distinction between metals and insulators. "Type B" modeling could also play an important role in biology. In "Type B" modeling, we set aside detailed analyses of biochemical networks in real organisms, and pursue instead simplified models in an attempt to elucidate or discover important biological principles. For example, Leibler and Barkai have produced a pared down model of chemotaxis in E-coli [3]. Although their model has limited quantitative predictive value, they were able to demonstrate that this model could be designed to display the interesting feature of biological "robustness", in this case an insensitivity of the "adaptability" (to changes in ligand concentrations) of flagellar rotation rates to major variations in rate constants and/or the concentration of important protein intermediates. Biochemical networks with linkages that display "robustness" may confer an evolutionary advantage as rate constants change due to genetic drift. See Sections 3.3 and 4.2 for detailed discussions of Type A and Type B modeling respectively.

The Barkai-Leibler work suggested important conceptual follow-up experiments and provided an idea which could reappear in other biochemical networks or organisms. Unfortunately, others have used the "robustness" concept indiscriminately to excuse the lack of knowledge of rate constants in more quantitative Type A models! Robustness, if it does apppear in a given biological network, can actually be accompanied by extreme *sensitivity* of other properties to parameter changes [3,4]. For example, the adaption time to changes in nutrient concentration in the Barkai-Leibler model is quite sensitive to random factors of two changes in kinetic coefficients, even though other important parameters are not (see Section 4.3). Although "robustness" may appear in a real biological organism whose biochemical "circuits" have the appropriate topology, it is by no means guaranteed that an approximate model constructed from imperfectly known rate constants will display this property. Indeed, there is at least one example of a Type A model of chemotaxis that failed to uncover the robustness property using a set of equations constructed with the best available experimental data [4].

"Type B" modeling has also been used to guide the top down synthesis of a switch involving a genetic network [5]. A pair of repressor genes, together with their promotors, was inserted into a plasmid appropriate for E-coli bacteria. The rate constants were then tuned (by varying that part of the repressor gene encoding for a ribosomal binding site) to produce a biological analogue of a "flip-flop" circuit in digital electronics [6]. Although the switching time in response to external inducer signals is slow (of order hours), decades of work on genetic engineering allows relatively straightforward tuning of rate constants in genetic networks of this kind. Such fine tuning is not as easy for the protein networks in, e.g., conventional metabolic pathways.

Using either Type A or Type B modeling, one could imagine biological research laboratories using computer simulations to test for the most promising avenues of research before going to a wet lab. In an ideal world, the experiments would lead to changes in the model that would, in turn, lead to more experiments. The possibility of a real ongoing dialog of this kind between theoretical modeling and experiments in the biological world is an exciting one.

In Section 2, we discuss the role of models in science and special problems associated with their application to biology. In Section 3, we describe attempts at quantitative "Type A" modeling, with an emphasis on models of Mycoplasm genitalium and of bacterial chemotaxis in E-coli. "Type B" modeling is discussed in Section 4, starting with the work of Hodgkin and Huxley on excitable biological media and moving on to simplified models of bacterial chemotaxis and of a biological switch in a genetic network. The concluding part of this section describes an interesting interpolation of an analog electronic circuit model of a lobster nerve cell into a network of real nerve cells. This report concludes with eight specific recommendations and conclusions in Section 5. It seems clear that biological models of cells will eventually need to deal with spatial variations in the constituents of various biochemical networks, even in prokayotes. Our recommendations include undertaking a survey of experimental techniques which could illuminate spatial and temporal variations with subcellular resolution.

References

[1] M. Tomita et. al., *Bioinformatics* **15**, 72 (1999)

[2] L. H. Hartwell, J. J. Hopfield, S. Leibler and A. W. Murray, *Nature* **402**, C47 (1999).

[3] N. Barkai and S. Leibler, *Nature* **387**, 913 (1997).

[4] T.-M. Yi, Y. Huang, M. I. Simon and J. Doyle, *PNAS* **97**, 4649 (2000).

[5] T. S. Gardner, C. R. Cantor and J. J. Collins, *Nature* **403**, 339 (2000).

[6] P. Horowitz and W. Hill, *The Art of Electronics* (Cambridge University Press, Cambridge 1990), Chapter 8.

[7] A. Szucs et. al., *Computational Neuroscience* **11**, 563 (2000).

[8] L. Shapiro and R. Losick, *Cell* **100**, 89 (2000).

2 BIOLOGICAL MODELS: WHAT, WHY and HOW?

In this section, we provide an overall rationale for the use of mathematical and computational models in biological systems. For several reasons, this concept is widely perceived as an idea whose time has come; these reasons include the vast flood of data being generated in a new generation of biological experiments, a general feeling that one will be able to get more useful information out of these data through quantitative approaches, and the perception that computational methods are lending useful insights into other complex systems, ranging from climate prediction to materials science.

Before attempting to define the role of a mathematical model, we should acknowledge that biologists have been using models for a very long time. Most often, however, these models are explicitly non-mathematical and typically amount to a set of connections that attempts to summarize the logical structure of a biological system. For example, Figure 2-1 taken from the work of Loomis and collaborators on the genetic network underlying development in the soil amoeba Dictyostelium discoideum is of this type. In this network "model", the pathway by which the receiving a signal (in the form of the chemical cAMP) leads to altered gene expression is laid out in logical, but not in mathematical/computational form; this model can be used for qualitative predictions but is clearly not capable of any quantitative conclusions.

So, what do *we* mean by a model? For physical science, a model is a set of mathematical relations that take some input variables, define a calculation to be performed (depending explicitly upon some model parameters) and generates some number of output variables. A model is not merely the

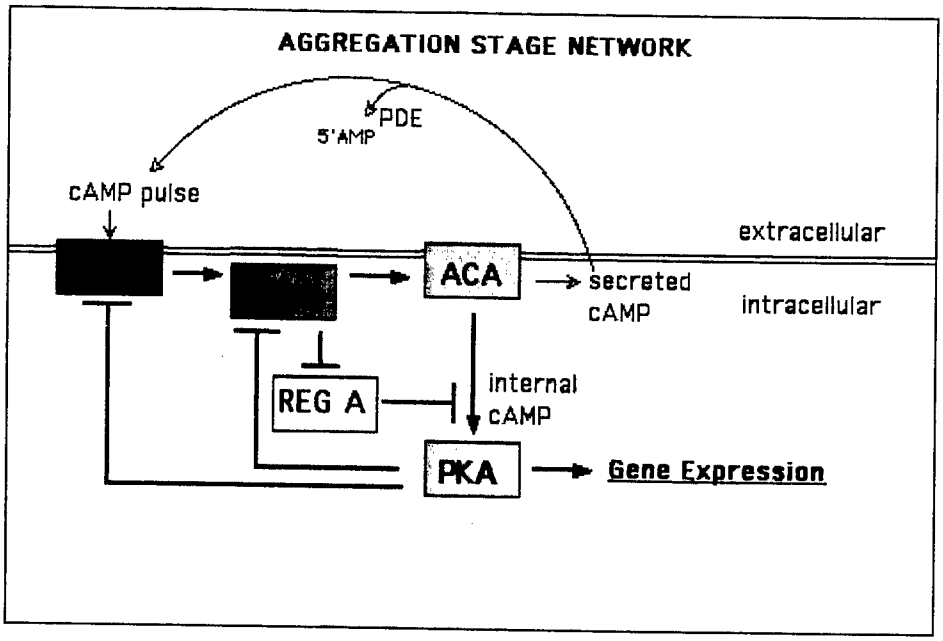

Figure 1:

summary of a given set of experiments in shorthand form. Instead, it represents an extrapolation (based both on data and on an intuitive picture of the involved mechanism) from existing information to unknown realms of input space and parameter space. Since a model is not just a re-writing of the data, it can be – and often is – dead wrong. Thus, the useful models are the ones that have undergone repeated cycles of having their predictions compared to new experimental data and still remain valid, albeit with inevitable modifications having occurred in the validation process.

A model as we have defined it is necessarily quantitative. Whatever the form of the output variables (Boolean, real numbers, probability distributions for stochastic systems etc.), these are uniquely determined by computing with the model. It may not be easy to do the necessary calculations as, for example, in the Navier-Stokes equation model for turbulent fluid flow, but in principle these can be carried out with zero uncertainty. (There is a caveat that a sensible model should have a match between the output and the

dynamics. For example, it makes no sense to claim that the Navier-Stokes equations will predict the velocity at some specific time/space point, as this is so sensitive to every last detail that it is a useless exercise; instead, one should define the output as various correlation functions, probability distributions etc., which are computable.) This caveat accepted, we will assume that our model makes quantitative predictions for the system under study. But, this does not mean that we expect these quantitative predictions to agree to arbitrary accuracy with experiments performed on the actual system. That is because all models are, by their nature, incomplete. Even the hallowed incompressible flow equations will not agree with data from fluid experiments to arbitrary accuracy; for example, there are clearly corrections due to the finite compressibility of any real fluid. Any model should therefore come complete with an understanding of how much the left-out ingredients will affect the model predictions. In some cases (such as the aforementioned flow case), this will be calculably negligible – these are of course models that physicists like best, and most of the familiar physical models (Navier-Stokes equation, Schroedinger equation, Maxwell's equations) are obviously of this type. In almost all complex systems, however, this will never be completely possible. Nevertheless, certain classes of models try to put in all the interacting pieces of the system and aim for this physics-style reductionism, recognizing of course that the results at any stage might only be semi-quantitatively accurate. In contrast, other models aim to get at the basic mechanisms with as simple a model as possible, and concede at the outset that only qualitative conclusions should be taken seriously. We will refer to these extremes as type A and type B models respectively, fully cognizant of the fact that there is no sharp dividing line.

Let us give some examples of these classes. In the model A genus are detailed climate models, ab initio band structure calculations, all-atom molecular dynamics simulations of biopolymer dynamics and large-scale computations of the evolution of astrophysical structure. These models explicitly attempt to get enough of the details right so as to render their results quantitatively reliable, with some implied accuracy level. As we will flesh out in more detail in the next section, there is some biological modeling that falls naturally into this class. This includes work on the E-cell simulation of Mycoplasma genitalium, the lysis-lysogeny switch in phage λ, and the BCT approach to E. coli chemotaxis. In this report, our focus is on using genomic and proteomic data to build models of cellular function; it is clear that these cell biology models will require massive complexity to reach even limited quantitative accuracy.

What use is a biological type A model? Essentially, one can substitute the model for the system and carry out computational experiments that might be impossible (or more difficult/costly etc.) to perform on the wet ware itself. This capability could have a large payoff; just imagine being able to design antibiotics without having to grow bacterial cultures in the lab. Even before complete replacement, type A models could guide experiments to the most effective regimes of parameter space; i.e., testing only a few drugs on bacterial cultures rather than having to do a full screening for every new hypothesis. Also, type A models can often be a stimulus for scientific advances, as a working model is prima facie evidence that the behavior of interest in some biological systems (say the ability of E. coli to detect chemical gradients on a wide range of background concentration levels) is indeed possible to accomplish with only the components used in the model

and no others. This, coupled with the ability to do extensive experiments, can rapidly spur increased understanding.

Perhaps the most important thing that needs to be done with a type A model is for it to be repeatedly confronted with experimental challenges. Biological systems are quite a bit better off in this respect than, say, climate or astrophysical models, since one can do a variety of critical experiments and try to test or refine the model.

This concept does not appear to be sufficiently ingrained in the biological modeling community. Again, an example from Dictyostelium can illustrate this point. Laub and Loomis have constructed a type A model of the cAMP signaling system, based on data collected in liquid test-tube cultures of cells undergoing starvation. As the model is based on the data (related to temporally oscillating levels of cAMP and other chemical species), it has been claimed that it does a good job of capturing the dynamics of the system. But, there have been no attempts to disprove the model by comparing its predictions to experiments that it was not explicitly designed to match. Hence, we have at present no reason to believe that the model would be a reasonable predictor of any cell response that was not put in by hand.

A similar problem underlies the use of learning algorithms to devise models whose only role is to fit experimental data – see, for example, models of gene expression in early Drosophila embryogenesis. Although exceptions exist, work to date can often be thought of as training a classifier with a training set of data, obtaining good fits to that data, but never testing whether the resultant network is effective at generalization.

At the other end of the modeling spectrum are type B models for which only the qualitative conclusions are claimed to have correspondence

with the actual system. Physicists sometimes refer to models of this type as "toy" models. Some examples of these include the models of nonlinear springs (which led to the notion of solitons), two component reaction-diffusion models of excitable media (such as the Oregonator model for the Belusov-Zhabotinsky reaction), lattice models of protein folding which utilize only two broad classes of amino acids (hydrophilic versus hydrophobic), the Burridge block-spring model of earthquakes, simplified traffic simulations and interacting walker models of pattern formation in nutrient-limited Bacillus colonies. The most successful of these lead to an appreciation of new mechanisms which turn out to be present in type A models as well, albeit hidden in the often more detailed equations. For example, solitons occur in type A models of oceanic internal waves and are quite important noise sources in the coastal region; study of these models has been greatly facilitated by a thorough understanding of similar mechanisms in simplified "toy" examples.

Given that type B models cannot quantitatively predict anything about a system, [not quite true; very simple models of phase transitions can get critical exponents right] why is it useful to construct them in a biological context? One possibility is that one might be able to directly engineer these models into cells and thereby create a useful new "designer biology". An example of this is the creation of a genetic toggle switch by starting from a model of the same and building it into E. coli (see Section 4.3). At a deeper level, though, there is another purpose. Type B models can represent an idealized picture of a mechanism that enables more understanding to emerge. Of course, "understanding" is a difficult concept to quantify and is highly subjective. But, we all would admit that the concept of "feedback from the output allows for perfect adaptation" as put forth by Barkai and Leibler to help explain E.coli chemotatic response characteristics, is most

clearly exemplified by the simple type B model they construct (see Section 4.2 for more details). Now, it may turn out that this concept is irrelevant to E. coli chemotaxis and hence their toy model offers no insight into how the system works and into how a type A model needs to be constrained. For example, decades worth of work on simple, "toy" reaction-diffusion models of morphogenesis, suggesting at their heart that Turing-type instabilities play the crucial role in organizing spatial patterns during biological development, now appear to be almost completely irrelevant. Here, basic biological facts, unknown at the time of Turing's work in 192, were left out of this example of Type B modeling. To reduce the chances of such pitfalls, it is again absolutely crucial that the model make predictions, even qualitative ones, that enable it to be falsified by experiment.

The major pitfall for type B predictions is losing track of the real system in favor of a mathematically well-defined game that becomes decoupled from experiment. The major pitfall for type A predictions is wallowing in all the details and losing track of the fact that models must be more than just curve-fits. In what follows, we will study in more detail some examples of biological models and try to understand the extent to which they have successfully escaped these traps and contributed to real scientific and technological progress.

Before concluding this section, we would like to point out why we feel that, in general, the introduction of serious modeling efforts for biological processes is a good idea, even from the point of view of experimentalists. A model can be useful as a methodology for assimilating disparate measurements and drawing links between them. In addition, models can provide useful frameworks for establishing a standard reporting protocol from disparate teams of investigators. For example, climate change models assimilate and integrate

inputs from oceanographers and atmospheric sciences, amongst others. Parameters that are necessary and critical for detailed understanding of a local phenomenon (for example, seasonal rainfall trends in a specific location or an understanding of tidal structures in a particular continental/oceanic boundary) are not necessarily the parameters of utmost importance to a general circulation model; the existence of the model often forces a re-prioritization for experimental work and spurs efforts to measure those quantities that are needed to critically test and thereby improve the model, as opposed to those which are crucial to describe more local, fine grain descriptions of the system or to improve an understanding of a more local phenomenon of interest to one subdiscipline or the other.

Similarly, the existence of a model often spurs consistency in form and format of data reporting, so that such data can be readily extracted for use by the modelers. This has the added benefit that it also makes results in a subdiscipline more accessible to those outside the field or entering it for the first time. In Boston, for example, major streets have no street signs because "everybody knows what those streets are"; similarly, a close-knit community of experienced investigators in a particular subdiscipline can each readily sort through a complex, qualitative experimental report in the field to find a few critical rate constants and perhaps implicit descriptions of the conditions under which they were measured and are valid. However, transferring this type of anecdotal information to an outsider is very difficult without a formal data reporting structure that is consistently used and applied by workers in the field. A good model can spur such systemization.

The development of widely accepted and used models in biology will face additional obstacles due to some special attributes of how biology has evolved as a discipline. First, biology is largely qualitative or semi-quantitative in

character. Secondly, there is a general disregard for theory and modeling in biology. Thirdly, biology is largely information-rich (details almost always matter). Fourth, biology itself is disparate in its culture, in that what is important to a cell biologist does not overlap much, in general, with what is important to a protein crystallographer which in turn does not overlap much with what is important to a geneticist.

A good model in biology thus should meet at least some of the following criteria: it should a) make a testable counterintuitive prediction, b) reveal a previously unknown general principle, c) allow experimentalists to manipulate a complex system in a predictable, useful fashion, and d) explain a heretofore poorly-understood phenomenon. Experimentalists will vote with their feet, so if a model makes an interesting and important prediction, experimentalists will test it. Conversely, if the model is unimportant, no one will pay attention. The best way to insure that a model has impact is to have the modelers involve experimentalists from the initiation of the effort, so that the modelers can work on problems that are important and of high priority to the experimental community. As the model evolves it will of course require additional sustained interactions between both communities in order to be validated, refined, and ultimately widely used.

3 TYPE A MODELING

3.1 E-Cell Program for Mycoplasma Genitalium

Mycoplasm genitalium is one of a family of exceptionally small (diameters as small as 0.2 μm) bacteria that lack a cell wall, and possess genomes only 1/2 to 1/5 the size of other free living bacteria. With its DNA consisting of 580k base pairs (M. genitalium was sequenced in the mid 1990's), the bacteria sustains itself with only about 500 genes. The Virtual Cell model inspired by M. genitalium [1] has 127 genes, which are "transcribed" (via rate equation modeling) to produce messenger RNA's which are then "translated" (again via rate equation modeling) by ribosomal RNA to make various proteins. Over 4000 molecular species involved in almost 500 chemical reactions are tracked by the program, which integrates a set of coupled nonlinear ordinary differential equations using Eulerian finite difference or 4th order Runge-Kutta techniques. Compared to real M. genitalium, the restricted gene set (127 "genes" in the model instead of $\sim$ 470 in the actual organism) does not allow for functions such as DNA replication and cell division. Instead, the network of chemical reactions defining the model takes glucose, fatty acids and glycerol and performs lipid synthesis and glycolysis, while carrying out the transcription, translation, and degradation of the proteins involved. The output of the glycolytic metabolic pathway is ATP, with lactate as a waste product.

We believe that there are considerable difficulties at present in obtaining reliable results for whole prokayotic cells, even for an abbreviated model of something as simple as M. genitalium. Knowledge of rate constants is a par-

ticularly severe problem. The rate constants chararacterizing a biochemical reaction in prokaryotes or eukaryotes appear as parameters governing a set of ordinary differential equations, as in

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} SE \xrightarrow{k_2} P + E,$$

where S, E and P are the concentrations of substrate, enzyme and product molecules respectively. SE is the concentration of the substrate-enzyme complex. Even in the well-studied lysis/lysogeny switch of the phage lambda which infects bacteria, the rate constants are only known to within a factor of 2, and often have to be inferred from a mixture of in vivo and in vitro experiments. Only three rate constants are required to parametrize the chemical reaction described above. However, *ten* unknown parameters are required to describe a more complicated bi-bi reversible enzymatic reaction with inhibitor and activator!

The rate constants k_{-1} and k_2 in the example cited above describe different modes of dissociation of the substrate-enzyme complex, and may have similar in vivo and in vitro values. In contrast, the important rate constant k_1, which describes the diffusive process by which substrate molecules and enzymes find each and dock, should depend sensitivly on the cytoplasmic environment. Indeed, a simple random walk argument shows that k_1 is given approximately by

$$\begin{aligned} k_1 &\simeq p a_e D_e \\ &\simeq p k_B T/(3\pi\eta), \end{aligned}$$

where $p < 1$ is the docking efficiency and a_e and D_e are the enzyme size and enzyme diffusion constant respectively. The last equality follows from the Stokes-Einstein equation, and shows that the product of the enzyme size and

diffusion constant drops out – what is crucial for k_1 is the *viscosity* η of the medium in which the chemical reaction takes place. This viscosity is a simple and relatively well understood parameter in many in vitro experiments. If one assumes a maximal docking efficiency $p = 1$ and inserts the viscosity of water, one obtains

$$k_1 = 10^8 - 10^9 / \text{ sec-M},$$

the well-known result for a "perfect" enzyme. However, the cytoplasm of a real cell is a complicated viscoelastic medium in which the lipid membrane and crumpled linear structures such as the bacterial chromosome can play an important role. It would not be at all surprising to have *orders of magnitude* differences in the effective in vitro and in vivo viscosities under some circumstances. Hence, lack of in vivo knowledge of k_1 for the many enzymatically catalyzed reactions in the E-cell model seems to us a particularly severe problem. The ratio of k_1 (in vivo) to k_1 (in vitro) should go as the ratio of the enzymatic diffusion constants in the two media. Thus, comparative measurements (using, say, green fluorescent protein tags) of D_e in vivo [2] and in vitro might allow in vitro measurements of k_1 to be converted to a number appropriate to an in vivo computer simulation. However, the viscoeleastic nature of the medium can still be a severe problem, especially in eukaryotes, where the apparent viscosity as inferred from particle diffusion is clearly a function of the particle size [3].

The difficulties sketched above would be far worse if one were to attempt a model of even more complex *eukaryotic* cells – the "circuit" that contains just the simple metabolic functions of eukaryotes (i.e., neglecting transcription, translation, regulation, etc.) is a complex network of about 500 nodes representing intermediates requiring knowledge of 2000 rate constants to specify the various enzymatically catalyzed links between these

nodes [4]. One of the defining characteristics of eurkaryotes is the presence of multiple compartments as well as linear/planar/bulk structures such as the nucleus, mitochondria, the endoplasmic reticulum, actin filaments, microtubules, the golgi apparatus, etc. Clearly, eukaryotes cannot be regarded as "well mixed chemical reactors", as may be the case for some prokaryotic functions.

With so many adjustable parameters in the form of poorly known rate constants (and enzymatic concentrations), it would not be surprising if simulations like E-cell could be fit to the exisiting experimental data. However, the choice of rate constants will certainly not be unique and one can question the value of this exercise for a system as complicated as an entire prokaryotic cell. As discussed elsewhere in this report, a handwaving appeal to the "robustness" of biochemical circuits is *not* an appropriate way out of this difficulty: "Robustness" of some quantities is accompanied by "sensitivity" in others. Here "robustness" of, say, the enzyme concentration in a particular biochemical circuit means that this concentration remains essentially unchanged in the presence of random changes in rate constants caused by genetic mutations. However, whether robustness exists at all in a particular organism is questionable in the absence of precise knowledge about the relevant biochemical circuits.

Although the multiplicity of poorly known parameters is the most severe difficulty with the E-cell simulation, there are other problems as well. A coupled system of, say, 4000 ordinary differential equations with ~ 10000 rate constants describing a simple prokaryotic cell should in principle encompass seven orders of magnitude in time scale, from milliseconds to hours. Under these circumstances, one expects many nested time scales. We can represent

such a system of differential equations schematically as

$$\begin{aligned} du/dt &= f(u,v,w,x,....), \\ \epsilon_1 dv/dt &= g(u,v,w,x,.), \\ \epsilon_2 dw/dt &= h(u,v,w,x,.), \\ &\vdots \end{aligned}$$

where $f, g, h, ...$ are order unity nonlinear functions of the reactant concentrations $u, v, w, x....$ Even for Michaelis-Menton kinetics, one typically finds (with appropriately rescaled variables) small parameters like ϵ_1 (e.g., the initial ratio of enzyme to substrate concentration) multiplying time derivatives. For many coupled reactions involving multiple time scales, one expects $\epsilon_1 \ll \epsilon_2 \ll ... \ll 1$. Special methods are required for integrating such "stiff" systems of differential equations [5]. Eulerian finite difference or 4th order Runge-Kutta techniques [1] are inadequate. This problem is not insurmountable, as systems equivalent to, say, 100,000 ordinary differential equations are routinely integrated by chemical engineers in efforts to model production lines in chemical factories. Another concern is that deterministic systems of ordinary differential equations must be modified when the number of molecules involved in a particular reaction is small. Fluctuations are then large and special stochastic simulation techniques are required [6].

References

[1] S. Tanida, K. Yugi, J. C. Venter and C. A. Hutchison III, *Bioinformatics* **15**, 72 (1999).

[2] M. B. Elowitz, N. G. Surette, P. E. Wolf, J. B. Stock and S. Leiber, J. Bacteriology **181**, 197 (1991).

[3] D. W. Provance, Jr, A. Mcdowall, M. Marko and K. Luby-Phelps, *Journal of Cell Science* **106**, 565 (1993).

[4] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts and J. D. Watson, *Molecular Biology of the Cell*, Chapter 2.

[5] W. Press et. al., *Numerical Recipes* (Cambridge University Press, Cambridge, 1985), Chapter 15.

[6] H. H. McAdams and A. Arkin, *Trends in Genetic Science* **15**, 65 (1999); *Ann. Rev. Biophy. Biomol. Struct.* **27**, 199 (1998).

3.2 Lysis/Lysogeny Switch

In a well-known paper McAdams and Shapiro [1] attempted to model the genetic switch by which phage lambda decides whether to lyse its host cell (lysis) or to instead incorporate its DNA into the bacterial genome (lysogeny). From our overall perspective, this system is a good choice for the type A paradigm; this is because many years of detailed biological legwork have succeeded in identifying all the components of the switch and have even provided some details regarding component kinetics. On the other hand, the numbers of molecules involved in some of the genetic control elements is sufficiently small that one needs to think about stochastic effects. This latter point has been emphasized by Arkin and McAdams [2], as discussed below.

The basic mechanism behind the switch depends on the competition between different promoters. The two basic genes, CRO and CI each try to lock in a transcription pattern which stabilizes its own production and suppresses the other's. If CRO "wins", the cell chooses the lysis pathway, whereas CI controls lysogeny. CI production is controlled partly by CII, where degradation is controlled by CIII; this process is affected by external conditions such as the state of nutrition and the number of infections per cell. These two competing pathways lead to bistability – either pathway can

stabilize itself and suppress its competition and it is just a matter of which chemical (CRO or CI) can accumulate more rapidly.

The basic workings of the switch had been elucidated by molecular biologists before the work of McAdams and Shapiro. Thereafter, these authors tried, in standard type A modeling style, to take a more or less known mechanism and transform it into a complete quantitatively predictive model. To do this, of course, requires knowledge of a whole variety of kinetic coefficients governing the basic reactions involved. There are at least 36 independent rate constants for the housekeeping/nongenomic reactions and the transcription/translation steps and around 10 more for the model of promoter control of transcription initiation. As is typical, one uses a variety of measurements plus a variety of desired network properties to guess at these parameters. Even in the simplest system, then, there is no real reason to trust the model quantitatively. This is especially true in that the typical tests of the model, the comparison of predicted time courses of important proteins with experimental measurements of the same are not carried out in a quantitative manner. Instead, one compares the trends produced in the model with the trends seen in the data. To the extent that these trends were already understood and the mechanism underlying them had been used to formulate the model in the first place, this is not sufficient to verify the modeling details.

As mentioned, the small numbers of some of the molecules involved as well as the random delays in the initiation of transcription suggest that a fully realistic model must be stochastic. This line of reasoning led to a detailed stochastic model that replaced the traditional reaction-diffusion equation for the various reaction steps by a directly simulated Markov process for (discrete) numbers of molecules. The simplest estimate of the noise in a reaction proceeding at rate R is that the fluctuations are proportional to $\sqrt{R}$; this

means that for either slow reactions or for small concentrations of reagents, noise is important. Not surprisingly, fluctuations will result in phenotypic diversity even for identical genomes and for identical initial conditions. In general, though, it will be very hard to disentangle intrinsic fluctuation effects from external variation in cell environment, changes in cell history etc. This has proven hard to accomplish even in the case of extremely well-controlled physical systems, and one should be wary of attributing phenotypic variation to stochastic effects without at least considering other possibilities.

In any event, stochastic models are much more complex and much more computationally expensive than are their deterministic counterparts. It is certainly interesting to use these models to estimate how much noise is actually present in the switching process. This probably could be done in a much simpler model unless one suspects that somehow the complexity of the actual phage switch arose partly as a way of compensating for having to work with noisy components. This interesting idea, unfortunately, rarely goes past the idea stage, possibly because the needed computations are still very expensive. Perhaps this is really a task for a type B model in which one can directly phrase the issue of how redundancy increases fidelity even with fluctuations, without getting trapped in a morass of irrelevant additional details.

In summary, then, the work of McAdams and collaborators represents a well-reasoned effect to formulate a type A model for the lysis-lysogeny switch. Their work is serious, their choice of system good and their rate of progress above almost all other similar efforts. So far, however, no great insights have been forthcoming, and, to the best of our knowledge, no counter intuitive predictions have been made and subsequently tested experimentally. It remains to be seen if the model, as presented formulated, is capable of quantitative

predictions in new experimental circumstances without additional fine tuning of poorly constrained input parameters.

References

[1] H. H. McAdams and L. Shapiro, *Science* **269**, 650 (1995).

[2] H. H. McAdams and A. Arkin, *Ann. Rev. Biophys. Biomol. Struct.* **27**, 199 (1998).

3.3 Bacterial Chemotaxis and the BCT Model

The best-studied biological organism is the bacterium *E. coli*, and the best-studied sensory system is bacterial chemotaxis. *E. coli* are attracted to a wide variety of simple chemicals, such as sugars, amino acids, and peptides, which serve as sources of food. These compounds need not be metabolized (nor even internalized) to be sensed. Bacteria are also sensitive to such things as temperature, oxygen, and pH. Moreover, they are repelled by potentially toxic compounds, such as certain metal ions. When confronted with a non-uniform distribution of a chemical that is sensed as an attractant (or repellent) in its environment, a bacterium will manage to swim up (or down) the chemical gradient. This chemotactic behavior is remarkably efficient, as it must be: bacteria need to outrun the diffusion of the very chemicals they sense, if they are to be captured as food sources, or to outpace the diffusion of toxic repellents. The physics of this problem is well understood (Berg & Purcell, 1977).

Bacteria swim by rotating their flagellar filaments, which are individually powered at the base by a nanoscale electric motor driven by a flux

of protons. A single cell of *E. coli* typically bears 8 flagella arising from random points on the cell surface, which are swept together to form a rotating bundle that pushes the entire cell forward. This is called a *run*, and it typically lasts around 1 second. During a run, the flagellar filaments all rotate in the counterclockwise (CCW) direction, as seen looking down the bundle towards the cell body. They spin rapidly, at speeds around 100-300 Hz. When one or more of the flagella reverses direction and rotates clockwise (CW), the bundle is observed to fly apart and the cell undergoes a *tumble*, randomizing its orientation. Tumbles typically last about 0.1 second. Both runs and tumbles are random processes, and their durations are governed by simple exponential distributions (Poisson statistics). In an isotropic environment, a cell of *E. coli* swims in a random walk consisting of alternating runs and tumbles. When placed in a gradient of attractant, cells modify their swimming by lengthening those runs that have a component in the favorable direction of the gradient. Runs that have an unfavorable component are unchanged, and governed by the usual baseline stochastic behavior. The same is true for tumbles. Overall, this results in a *biased random walk*, which can be modeled as diffusion with drift. For actual bacteria, the drift rate is roughly 10% of the raw swimming speed. The latter can be as high as $\sim$35 μm/s (about 20 body-lengths per second). Bacteria sense spatial chemical gradients by swimming through these while monitoring the temporal rate of change of specialized chemoreceptors, called transducers, that bind to one or more kinds of ligand (e.g., an amino acide like aspartate). A cell of *E. coli* carries about 5,000 transducers in all, consisting of five basic varieties, each capable of responding to a given subset of all sensed chemicals.

Bacterial sensory behavior can thus be viewed as an input-output system in which the input is the time-varying chemical concentration in the

cell's immediate environment, and the output is an internal signal that modifies the chance that the rotary motor will switch back and forth between the CW and CCW directions. Bacteria therefore take a relatively noisy, analog variable (chemical concentration) and transduce it into a stochastic, binary variable (CW/CCW rotation). Bacteria are so sensitive that they can respond to a change in occupancy of only a single chemoreceptor (out of a thousand) over the time course of single run ($\sim$ 1 s): this amounts to extraordinary sensitivity. The sensory pathway of *E. coli* is quite evolved, and displays many of the same hallmarks of analogous sensory modalities in higher organisms. In addition to its sensitivity, it has a large dynamic range, and can respond to chemical concentrations ranging over nearly five orders of magnitude. It also *adapts*, so that permanent changes in concentration produce only transient responses of the sensory system. Put differently, *E. coli* can respond to the rate of change of chemical concentration (equivalent to a concentration gradient), and not to the concentration *per se*.

Introduction to the Chemotaxis Pathway

How are *E. coli* able to do this? The sensory transduction system in *E. coli* is remarkably simple: it consists only of the five kinds of chemical transducer (named *tar*, *tsr*, *trg*, *tap*, and *aer*) plus exactly six kinds of *che* (chemotaxis) proteins, so named because a loss of any one of these genes leads to a generally-nonchemotactic phenotype. The genes are named *che* A, B, R, W, Y and Z. Together, the *che* genes form a compact signaling pathway, with both feedthrough and feedback, that acts stochastically to throw the flagellar motor switch, which in turn consists of a complex of three flagellar genes, *fli* G, M, and N, which form a portion of the motor itself. In its simplist form, the entire pathway is therefore comprised of **eight** distinct components: a transducer, 6 *che* proteins, and a switch. However, the appar-

ently simple composition of this system belies an astonishing sophistication and complexity.

A detailed characterization of this pathway has been gained by about 35 years of hard work in this system by many scientists, dating back to the seminal work of Julius Adler (Wisconsin) in the mid-1960s. Thanks to the completion of the *E. coli* genome project, all chemotaxis and flagellar genes were exhaustively identified and sequenced (in fact, all but one gene, coding for the *aer* oxygen sensor, had been found by other means, long before the completion of the bacterial genome). A multitude of mutants is presently available from any of several dozen labs working worldwide on chemotaxis. Protein structures have now been solved for at least three of the chemotaxis-pathway polypeptides, cheA, cheR, and cheY, and also for the transducer, tar. In addition to the genetics, a great deal is known about the biochemistry and physiology of *E. coli* chemotaxis. The study of *E. coli* physiology has been advanced by the development of the *tethered cell* assay. In this assay, a bacterial cell is affixed to the glass surface of a microscope coverslip by means of a single flagellar filament (the other 6-7 flagella are first broken off or otherwise removed). When the cell attempts to rotate its attached flagellum, the cell body is forced instead to turn round and round. Using the tethered cell assay, it is possible to monitor the output of a single flagellar motor on a cell. It is also possible to challenge the cell with gradients of chemicals and see the effect of these on motor directional switching, and readily to characterize mutants of various sorts. In this fashion, the 'black box' response (system function) of *E. coli* was well characterized in the 1980's, so that a quantitative mapping of time-varying receptor occupancy to motor switching probability is known.

Chemotaxis Biochemistry

To begin the process of modeling bacterial chemotaxis, it is first necessary – but not sufficient – to understand the overall connectivity of the chemotaxis biochemical pathway. When a simple chemical (say, the amino acid aspartate) binds to or unbinds from a transmembrane transducer (say, tar), this change in state triggers a series of biochemical events. The transducer is bound on the cellular side of the membrane to two other proteins, the products of the *che*W and *che*A genes, in a ternary complex. (A complication: often, this ternary complex is itself dimerized, or present as an even higher-order multimer, so that there are multiple copies of tar, cheW, and cheA.) The cheW protein serves as an adapter to hold cheA. CheA is a special type of histidine kinase, and it is capable of self-phosphorylation at a rate that depends on several things, including, importantly, the receptor occupancy of the transducer. Once auto-phosphorylated, cheA-P can transfer its phosphate to one of two other proteins: cheY or cheB. CheY is a small soluble protein ($\sim$ 12,000MW) that can diffuse across the volume of the bacterial cell in milliseconds. In its phosphorylated form, cheY-P can bind to the flagellar motor switch and induce CW (tumble direction) rotation: it therefore acts as a "tumble signal". The cheA-cheY phospho-relay combination acts as a feedthrough, or excitation, pathway. Bacterial excitation is fast, and occurs within a fraction of a second (< 100 ms) of the time a sensed chemical binds or unbinds from the receptor. CheY-induced excitation is not permanent: it is eventually turned off by the activity of a specific phosphatase, cheZ, which returns cheY-P (signaling) to the cheY (non-signaling) state. But cheA can also transfer its phosphate to cheB. CheB-P acts more slowly, over a period

of about four seconds, to modulate the baseline activity of cheA-P. This constitutes the feedback, or *adaptation*, pathway. It works in a somewhat more complicated fashion, as follows.

The activity of cheA bound to its transducer not only depends upon the amount of chemical ligand bound, but also upon the methylation state of the transducers, which are also known as "methyl-accepting chemotaxis proteins", or MCPs. Transducers have at least four sites (glutamine residues) that can be reversibly methylated by an enzyme that is the product of the *che*R gene (a methyl transferase). CheR activity does not appear to be regulated. But the steady-state level of methylation of any given transducer depends on the balance between its methylation rate and demethylation rate. Methyl groups are removed by the product of the *che*B gene (a methyl esterase). When phosphorylated, cheB-P becomes a more active esterase than unmodified cheB, so that methyl groups are removed more efficiently and the steady-state level of the transducer methylation falls. Thus, feedback generated by cheA-P, acting through cheB-P, leads to the removal of methyl groups from the transducer. This, in turn, lowers the autophosphorylation rate of cheA in the transducer-cheW-cheA complex, thereby *turning off the phosphorylation signal initiated by ligand binding itself.* The level of methylation therefore acts as a kind of "scratchpad memory" for chemicals, reflecting inside the cell what's happening on the outside. The higher the level of chemical in the environment, the higher the level of methylation.

To recap: the binding of chemical ligand leads to autophosphorylation of cheA. CheA-P sends its fast excitation signal immediately to the motor, via a cheY-P phospho-relay. CheY-P is then eventually degraded by cheZ. CheA-P also sends a slower feedback signal to the trandsucer, via a cheB-P phospho-relay, which removes methyl groups from the transducer and lowers

the activity of the attached cheA kinase. One last complication is that the whole chemotaxis pathway is wired up for negative regulation, in the sense that it is the *loss* of bound ligand that leads to an increase in cheA activity and thereby an increase in tumbling. The suppression of tumbles during runs that have a component up the gradient of an attractant leads to chemotaxis, as stated before. Put differently, the system seems to act as though it responds to the loss of attractant by tumbling, rather than gain of attractant by swimming smoothly. Incidentally, the bacterial response to the addition of a chemical repellent is identical to the response to the loss of an attractant (and vice versa).

Complications

The description presented above was a simplified version of the full biochemical pathway, as it is currently understood. There are quite a few additional phenomena known to complicate matters. First, the ternary transducer-cheW-cheA complex can form dimers and multimers, with varying numbers of protein constituents and therefore many possible levels of activity. Second, the activity of these complexes may depend in complicated ways on the ligand occupancy. Third, the number of glutamate residues that are potentially methylated on the transducers may vary. Fourth, the enzyme cheB fulfills another role in addition to its function as a methyesterase: it can also transform glutamine residues on unprocessed transducers thereby rendering them capable of methylation by cheR. Fifth, there are actually two forms of cheA polypeptide made in bacterial cells, one somewhat longer than the other, known as cheAL and cheAS. The biochemical roles served by these two different forms are not known. Sixth, the activity of cheZ may be regulated in ways that are incompletely characterized, for the present. Seventh, there may be interactions among the five different classes of transducer, known

as "crosstalk". Finally, there are several unresolved questions surrounding motor switch cooperativity and mechanical feedback. *Therefore, even in a pathway of just 8 components, there are a multitude of complications.*

The BCT Program

The bacterial chemotaxis (BCT) Program evolved from initial attempts by Dennis Bray and colleagues (Cambridge University, UK) to develop a more-or-less complete, computer-based model of the bacterial chemotaxis signal transduction pathway. This pathway represented a likely choice, given the tremendous body of literature in this field, consisting of well over 200 papers, and the fact that all protein components were identified, all genes were sequenced, and much of the basic biochemistry was established. Early attempts at simulation were based on solving numerically a series of coupled, deterministic rate equations which represented a subset of the known chemical reactions. The original efforts were written in a version of BASIC and employed Runge-Kutta integration methods, but produced unsatisfactory results for a variety of reasons. The present version is written in C++ and no longer solves deterministic rate equations. Instead, it is based on a kernel called StochSim that represents each molecule in the simulation as a separate program object: this proved feasible because there are only on the order of a few thousand molecules involved in the complete chemotaxis process. The StochSim engine represents each reaction as a stochastic (MonteCarlo) process, and therefore (at least in principle) faithfully captures the chemistry of the system, even when the number of molecules involved is small.

BCT, as currently constituted, does not attempt to simulate the complete bacterial chemotaxis pathway: it represents a 'stripped-down' version. For simplicity, BCT models only a single type of transducer protein, tar, the receptor responsible for binding aspartate. The solitary input to the system is

therefore the time-varying concentration of aspartate in the virtual medium, which is assumed to be in instantaneous equilibrium with the tar transducer. (By omitting receptors of the tsr, trg, tap, and aer classes, BCT bypasses a number of currently unresolved experimental issues surrounding receptor 'balance' and crosstalk among parallel signaling pathways.) BCT models all six of the che proteins, as well as all phosphorylated and methylated derivatives of these. The running output of the BCT model is the instantaneous probability that a single flagellar motor will spin either CW or CCW: this is computed directly, by using an assumed nonlinear functional form, from the concentration of CheY-P in the simulation. The BCT chemical "computer" therefore maps an external aspartate concentration to an intracellular cheY-P concentration.

Despite the small number of enzymes involved in the chemotaxis pathway, there is a 'combinatorial explosion' of possible reactions and corresponding rates. In BCT, for example, over 60 different chemical reactions are directly simulated. This proved necessary because the minimal signaling complex that leads to cheA phosphorylation consists of at least three proteins (tar, cheW, and cheA), which must come together to form a ternary complex before binding the attractant, aspartate. Because these three proteins can associate together in various alternative sequences, and because of the propensity of tar to dimerize (creating tar-cheA-cheW hexamers), and because there exist multiple methylated forms of tar (every tar receptor carries from 0 to 4 methyl groups), each with different activity, enumeration of all possibilities leads to a formidable number of equations.

BCT Model Evaluation

The BCT model can lay claim to a number of successes in modeling the chemotaxis system. In simulations of 65 known deletion and/or over-

expression mutants in the chemotaxis pathway, BCT was able to 'postdict' 60 of these correctly, as well as to predict the behavior of a number of previously unexamined genotypes. Robert Bourret (Univ. North Carolina) is currently attempting the construction of some of the latter strains, and so more stringent tests should soon become available. The $\sim$ 10% of all strains that were incorrectly predicted were attributed, in part, to incompleteness of the model, e.g., from effects arising from other transducer pathways (Levin et al., 1999).

Here are some future challenges for the BCT system:

- *Selected rate constants don't agree with experiment in a significant number of cases:*

  Many of the $\sim$60 rate constants selected as input parameters to the BCT model are set, where possible, to actual values reported in the literature. The experimental numbers typically differ from lab to lab (by factors of two to five; more in some cases), necessitating the choice of a compromise value. However, in a number of instances, BCT parameter values used for 'successful' runs differed by several orders of magnitude from experimental values. This raises the question of whether the relevant experimental values were incorrect or inappropriate (e.g., that the circumstances of the measurement in vitro did not reflect the activity in the cell in vivo), or whether the BCT simulation explores the wrong region of parameter space.

- *Modeled cells don't adapt fully:*

  To get complete adaptation in the model to permanent changes in the ambient concentration of aspartate, certain rate constants must be adjusted and maintained at rather precise values, a process referred to

as "fine tuning" – something that is thought to be unachievable in real-world cells. Barkai and Leibler (1997) have argued that reaction pathways often need to be robust against perturbations in concentrations and reaction rates, and proposed a general mechanism to accomplish this using a reaction circuit employing selective feedback. If the Barkai and Leibler mechanism is incorporated *ad hoc* into BCT, it does achieve more-or-less complete adaptation. However, doing so requires setting certain rate constants to zero that were thought to be otherwise, i.e., altering the connectivity of the reaction pathways. So there are unresolved issues here that remain to be addressed.

- *The bacterial impulse response function has the wrong shape:*

  This is somewhat related to the previous discussion. Again, unless the fine-tuning of rates constants is invoked, the overall kinetics of the BCT response to an impulse of aspartate (this is the *system response function*, in the language of engineering) does not correctly reproduce the experimentally-determined function.

- *Modeled cells don't respond over the full dynamic range:*

  Bacteria can respond to chemical concentrations over a dynamic range of roughly five orders of magnitude. However, computer simulations fail to achieve responsiveness over this range. A related issue is the astonishing sensitivity of the chemotactic apparatus (see below).

- *BCT doesn't get 3D chemotaxis right:*

  The BCT simulation is, by construction, incomplete, because it only computes chemotaxis to the level of the cheY-P concentration. With some additional assumptions, one can map cheY-P concentration onto the probability that a single motor will spin CW or CCW. However, it

is not understood how the rotary behavior of a single flagellum maps onto run/tumble behavior of a bundle of flagella powering a swimming cell. For example, the average CW period of an unstimulated motor in a tethered cell is about $\sim$1 s, whereas free-swimming cells tumble for just $\sim$0.1 s, on average. The source of this discrepancy is not well established, and may be due to mechanical interactions among the flagella within a bundle. It may also be due to hydrodynamic forces, mechanical feedback, the differences in load between tethered and free-swimming cells, the statistics of reversals, etc.

A second difficulty is the experimental asymmetry of responses in gradients. Although real cells running up a gradient of attractant lengthen their runs, cells running down the same gradient revert to baseline behavior, and do not shorten their runs. This 'rectification' phenomenon is also reflected in studies of tethered cells in gradients, which show differing thresholds in responses to up- and down-ramps of attractant. The BCT simulation has not been reported to predict either of these phenomena.

Of course, these are not failings of the model *per se*, which is known to be incomplete, but they prevent detailed simulations of the BCT type from predicting quantitatively the behavior of actual bacteria swimming in real-world gradients.

- *Modeled cells don't manifest the observed sensitivity to chemical gradients:*

  In addition to a large dynamic range, it has been shown experimentally that *E. coli* can respond to the change in occupancy of a single receptor over the course of a single run. This poses a problem for simulations, which thus far do not have enough intrinsic gain to display

single-receptor sensitivity. The gain problem, however, has stimulated additonal thinking and a healthy dialog between modeling and experiment. It can be fixed, in principle, but most attempts to do so sacrifice dynamic range and are therefore unsatisfactory. It has been suggested that models may need a different kind of amplification. Among the proposals currently under consideration are variations of the 'zero-order ultrasensitivity' mechanism of Goldbeter & Koshland (1981), or certain types of motor cooperativity (H.C. Berg and S. Leibler groups). A recent proposal by the Cambridge BCT group (Bray, Levin, Morton-Firth, 1998) has advanced the idea that chemoreceptors can form heterogeneous clusters of varying size, ranging from a few chemoreceptors up to several thousand. If receptor clusters are capable of signaling in a cooperative fashion, this mechanism is purported to show both large gain and dynamic range. It has not yet been incorporated as an integral part of BCT.

Conclusion

The BCT program is a casebook example of the astounding complexity that can quickly develop from attempts to model "simple" examples of biological networks. Even with just 6 pathway enzymes, there are well over 60 reaction rates and about 10 concentrations that need to be considered, leading to an enormous number of degrees of freedom. Unless a number of additional features are imposed on the underlying model *ad hoc*, such as the 'robust' feedback control of adaptation, chemoreceptor clustering and cooperative signaling, motor switch cooperativity and rectification, etc., it is difficult to reconcile detailed simulations with all the known features of chemotaxis. Its always possible of course, that important connections in the biochemical network of chemotaxis are left out, even in a detailed Type

A model. The bacterial chemotaxis system supplies an excellent test case, however, since it is one of the few systems where *quantitative*, as well as qualitative, data are available for the physiology, where the genetics is completely established, and where the biochemistry is fairly complete. In light of the experience already gleaned here, it seems reasonable to suppose that any difficulties in modeling this superbly well-established system will be experienced in even greater measure when confronting metabolic pathways, developmental pathways, gene regulatory pathways, sensory modalities, and other systems of interest in higher organisms.

References

[1] H.C. Berg & E.M. Purcell, The physics of chemoreception, *Biophys. J.* **30**, 193-219 (1977).

[2] C. J. Morton-Firth & D. Bray, Predicting temporal fluctuations in an intracellular signalling pathway, *J. Theor. Biol.* **192**, 117-128 (1998).

[3] C. J. Morton-Firth & D. Bray, Predicting temporal fluctuations in an intracellular signalling pathway, *J. of Theor. Biol.* **192**, 117 (1998).

[4] M.D. Levin, C.J. Morton-Firth, W.N. Abouhamad, R.B. Bourret & D. Bray, Origins of individual swimming behaviour in bacteria. *Biophys. J.* **74**, 175-181 (1998).

[5] N. Barkai & S. Leibler, Robustness in simple biochemical networks. *Nature* **387**, 913-917 (1997).

[6] D. Bray, M.D. Levin, C.J Morton-Firth, Receptor clustering: a cellular mechanism to control sensitivity. *Nature* **383**, 85-88 (1998).

4 TYPE B MODELING

4.1 The Hodgkin-Huxley Model

One of the best known examples of biological modeling is afforded by the Nobel Prize winning work of Hodgkin and Huxley (HH) on action potential propagation down the squid axon. The biophysics of this problem concerns the fact that the conductance of the cell membrane is governed by voltage-sensitive ion channels. Typically, the concentration of sodium is much higher outside the cell, the reverse being true for potassium; the resting potential is roughly -80 mV, set mainly by the Nernst equilibrium voltage for potassium. When the membrane is depolarized to, say, -40 mV, sodium channels open and, via the above concentration gradient, there is a large influx of Na ions leading to further depolarization. The channel kinetics lead to a time scale of milliseconds for this voltage change. Afterwards, the sodium channels close, the potassium ones open and the resting voltage is restored.

We regard the Hodgkin Huxley work as an example of "Type B" modeling, in part because a number of important physical processes involved in nerve conduction were simply not known at the time of its creation in 1952.[1] This model, nevertheless, gets the basic physics and biochemistry right and has stood the test of time. It can also be regarded as a bridge between simpler, more analytically tractable Type B models and Type A versions which attempt to add more details (see below).

The key to action potential propagation is that neighboring parts of the membrane are coupled. The voltage at one point will cause a current to

flow to neighboring sites, leading to their excitation. This system was then modeled as a (one-dimensional) partial differential equation for the voltage which is coupled to several (for HH-three) ordinary differential equations governing channel response and resultant conductivity charges. A Type A version of this model would work up from a molecular scale description to create an ab initio model of how this physics occurs. In fact, the molecules involved (the ion channels) were not even identified at the time of the HH work, and instead the currents were modeled phenomenologically. It was (and is) a great success story, in that the basic mechanism of membrane excitability has remained the key to electric wave propagation ever since, even as many molecular constituents were discovered experimentally and added to the repertoire of any given cell type generating an action potential.

It is interesting to note what has happened since the HH model of almost half-century vintage. One research direction has pushed this work further into the model B camp by finding more analytically tractable equations with the same basic excitability mechanism. The most notable of these is the Fitzhugh-Nagumo two component reaction-diffusion model [2] which serves as the simplest exemplar of nonlinear waves in excitable media. This further idealization has proven crucial in efforts to understand phenomena more complex than 1d waves, notably spiral waves in two dimensions and the possibility of spatio-temporal chaotic states.

Conversely, there has also been a push towards type A versions. For example, the Luo-Rudy model of electrical waves in heart tissue [3] tries to incorporate all available data on a multitude of ion channels as well as keep track of complex internal calcium dynamics, in an attempt to explain quantitatively details of the action potential. There is even a push to include the stochastic effects arising from channel openings and closing (see Section

3.2). The model suffers from all the usual type A problems; for example, data are not available for all channels in a specific organism and thus experimental findings from rabbit, dog and other organisms are blithely combined. Again, it is not clear if the attempted faithfulness to the molecular basis of ion conductivity is worth the price of having a large unwieldy set of equations of unknown reliability.

References

[1] A. L. Hodgkin and A. F. Huxley, *J. Physiol.* (London) **117**, 500 (1952).

[2] R. Fitzhugh, *Biophys. J.* **1**, 445 (1961); J. S. Nagumo, S. Arimoto and S. Yoshizawa, Proc. IRE **50**, 2061 (1962).

[3] C. Luo and Y. Rudy, *Circ. Res.* **68**, 1501 (1991).

4.2 Analysis of Protein Networks

Let us return to the E. coli chemotaxis system, which as discussed earlier, is one of the best characterized molecular systems in biology. Two papers [1,2] on this subject appeared at about the same time 3 years ago, both addressing the issue of "adaptation", i.e., bacteria's ability to respond *only* to transient changes in the level of external stimuli (see below). However, the objectives of these studies, as well as the methodology used and the conclusions reached, could hardly be more different. In the following, we contrast these two studies in some detail to illustrate the two different types of modeling and their respective strengths and weaknesses.

By "adaption", we mean an important aspect of the complex chemotaxis signalling pathway discussed in Section 3.3. Adaptation insures that bacteria such as E. coli are sensitive to *gradients* of nutrients like aspartate and not to absolute concentrations. When the food supply is constant in space and time, the flagellar motor "idles" with a mixture of runs and tumbles which mimics unbiased diffusion. If a bacterium is then subjected to a step function change in the background aspartate concentration l of, say, $\Delta l = \pm 1\mu$ M, the intervals between tumbles will temporarily lengthen or shorten in response to this sudden temporal change. This is the mechanism by which bacteria respond to small spatial gradients by sensing a temporal change in the concentration of attractant. However, after an *adaption time* τ, the bacteria μm adjusts to the new uniform concentration embodied in the step function and the motor reverts to idle. The *adaptation precision* is the accuracy with which a motor reverts back to its initial idle "speed" or switching rate. As discussed in Section 3.3, bacteria respond readily to nutrient gradients over many orders of magnitude of nutrient concentration. Adaptation precision allows this insensitivity to the background nutrient concentration and is a desirable property for any mathematical model of bacterial chemotaxis. The presence of aspartate gradients, a useful model should also lead to diffusion with drift.

The work of Spiro et al. [1] is the classic example of type-A modeling: as declared in their abstract, the study "incorporated recent biochemical data into a mathematical model that can reproduce many of the major features of the intracellular response, including the change in the level of chemotactic proteins to step and ramp stimuli such as those used in experimental protocols." Thus, the aim of modeling here is to *reproduce* experiments quantitatively, taking the existing biochemical data as constraints. In order to

accomplish this goal, Spiro et al. found that a minimal model must have three (methylation) states. And to achieve adaptation, the rate constants in the system need to be "tuned by trial and error". It turns out that despite 35 years of extensive biochemical studies of the chemotactic system, a great deal of freedom (i.e., unknowns) remains, and many aspects of the experiments can be reproduced by a sufficiently complicated model only after fine-tuning their parameters. The situation here is similar to another Type A theory of bacterial chemotaxis, the BCT model discussed extensively in Section 3.3.

Barkai and Leibler instead address "the issue of the sensitivity of the networks to variations in their biochemical parameters". They started from a *qualitative* demand on the system, that the ability of the cell to adapt to external enviroment should be "robust", e.g., insensitive to intracellular enzyme concentration, and *challenged* the models to reproduce this property.

Barkai and Leibler show how the *topology* of a biochemical network can insure robustness in the context of simplified Type B model. Their chemotaxis network tracks only 3 enzyme concentrations and requires only 9 rate constants as input parameters, in contrast to the 7 proteins and 50 rate constants used, for example, in the BCT model discussed in Section 3.3. Despite its simplicity, the model parameters can be chosen to mimic diffusion and drift in the presence of concentration gradients and to display adaptation to sudden changes in the background nutrient concentration. These authors then study the response of their system of differential equations to a series of random factor of two variations in the parameters. These variations, which simulate the effect of genetic mutations, lead to an ensemble of models defined by over 6,000 different parameter sets. Remarkably, the adaptation precision to sudden changes in the nutrient concentration remains extremely high for a very large fraction of these models. Robustness of the adaptation precision

to parameter changes arises from the network topology, and does *not* require additional fine tuning of parameters. In contrast, the adaption *time*, initially of order 10 minutes, varied greatly from 1 to 100 minutes over the space of 6,000 different models. Thus the adaptation time is *not* robust, in contrast to the adaptation precision. A chemotaxis network with the topology of the simplified Barkai-Leibler model might arise if evolutionary pressures favored developing biochemical networks with robustness in the adaptation *precision* and the adaptation *time* were less important. The simplicity of this model allows a clear identification of a simple biochemical network whose topology leads to robustness. The property of robustness, uncovered by this analysis, could be more general, with realizations in the biochemical circuits of other organisms.

The demand for robustness is reasonable to most biologists, as adaptation is a critical property for the survivial of bacteria. However, none of the existing quantitative models (e.g., refs [1],[3] and [4]) are robust in the sense defined by Barkai & Leibler. A detailed inspection of these models reveals that the connectivity of the networks in the models simply does not allow robustness. Barkai and Leibler demonstrate instead that a simple network with different topology (two internal states of the receptor-kinase complex, with demethylation acting only on those in the "activated" state) can attain robustness without any fine tuning. The essential ingredient of this model is the direct feedback of only the quantity being regulated, in this case, the fraction of activated receptor-kinase complex.

The Barkai-Leibler work is an example of Type B modeling. As discussed above, the goal is not to get all the quantitative details right, but rather, to demonstrate some qualitative principle, which can easily be missed from the detail-laden bottom-up approach typical of type-A modeling. Crit-

ics of the Barkai-Leibler model (and more generally type-B modeling) may justifiably charge that this approach lacks quantitative predictive value, leaves out the robustness of other important quantities, and relies on some hitherto unsubstantiated interaction pathway (in this case, the specificity of the demethylase on the hypothesized "activated" form of the receptor-kinase complex.) Nevertheless, the qualitative analysis of the Barkai-Leibler work, along with evidence from the follow-up experimental work supporting the hypothesis of the robustness of adaptaion [5], rules out practically all of the pre-existing quantitative models, suggests new experiments to probe the network topology, and perhaps most importantly, presents a plausible simple mode of feedback regulation which can occur in a much wider range of systems than the bacteria chemotactic system it is proposed for.[1] Thus, appropriate type-B modeling complements type-A modeling and can be very beneficial in getting the "big picture" right. Unfortunately for historical and social reasons, type-B modeling seems under-appreciated in the biology community.

It is important to point out what we view as strength in the Barkai-Leibler work is not so much the property of robustness itself. Rather, it is the recognition that certain properties of the system are robust while others are not. The opposite of robustness is "sensitivity", which may be as important (or even more important) to biology than robustness. [In the context of engineering, it is often recognized that as one makes certain aspects of a system more robust, other aspects become more sensitive [7].] A modeler ought to be able to identify in his model some components which are robust and others which are sensitive, and understand why they are so. These identifications then generate falsifiable predictions which can be tested by

[1]This particular mode of feedback regulation is in fact very well known in engineering and control theory [6].

experiments. Solid progress in knowledge will only come after close dialogue between modeling and experiments. The best way to facilitate such dialogue is to encourage both modeling and experimental efforts in the same group.

Short of direct synergy, it should be the responsibility of the modelers to make sufficiently clear predictions (e.g., which properties are sensitive and which are robust) to facilitate and stimulate new experiments.

4.3 Analysis of Gene Networks

An example of type-B modeling at the level of gene regulation is the recent study of the Drosophila segment polarity network by von Dassow et al. [8]. As in the Barkai-Leibler work discussed above, von Dassow et al. pointed out that the topology of the pre-existing network was insufficient to generate the desired output, in their case, the stable, periodic but asymmetric expression of the constituent genes wg and en. They noted that by adding a few more interactions (consistent with experimental facts) to the network, one can obtain the desired output with a wide range of parameters. They thus identified the segment polarity network as a robust developmental module.

While the von Dassow work is a rare example of type-B modeling at the *multicellular* gene network level (with spatial coupling), it falls short in terms of analysis, as the only "analysis" in the paper consisted of random sampling in the vast parameter space. Consequently, this work can only be regarded as suggestive of the existence of some robust module; they in fact never identified the underlying module. The failure to identify the underlying module and the sensitive nodes in this biochemical network makes the model lack predictive value and possible universal transportability to other related systems. Thus by the criterion we stated above for type-B models, the work

of von Dassow et al. represents an exciting beginning – future efforts to extract the underlying module and identify the robust/sensitive components are needed to turn it into a mature type-B model.[2]

One particular attractive feature of the gene network is that it is much more convenient to design and synthesize than the protein network – one can choose from the vast number of known DNA-binding proteins, and adjust various kinetic coefficients "simply" by modifying the binding sequences. Thus qualitative understanding gained from type-B can be tested in vivo using appropriately constructed gene networks.[3]

Gardner, Cantor and Collins [9] constructed a simple gene "network", a 2-state toggle switch similar to that controlling the lysis/lysogeny network discussed in Section 3.2. Their biochemical switch is the genetic analogue of a flip-flop circuit in digital electronics. Although the natural parameters of the genes inserted into E. coli led to a monostable steady state, bistability was achieved by fine tuning promoter sequences. The switching time was quite long, of order several hours. Nevertheless, one can imagine applications such as sensors and changing gene expression in plants in time of drought. Encouraged by their initial success, Canter et al. plan to go on with the multiplexing of toggle switches to design more complicated logic functions. An interesting theoretical question which arises in this context is the integrity of the computation carried out by a network of chemical reactions, which are subject to much stronger fluctuations than conventional logic circuits (e.g.,

[2]A brief inspection of the model indicates that the key component is a 3-state switch which remembers/rectifies a range of initial conditions. The periodicity and asymmetry of the segmentation patterns are put by hand into the initial conditions, i.e., they are fine tuned.

[3]For applications, the down side of gene network is its slow temporal response, order of minutes to hours, whereas the protein response time is much faster. It was suggested by Charles Zuker that gene networks may be used as storage devices which record various encounters by a cell.

sensitive temperature dependence of individual rate constants, stochastic effects in the presence of a small number of molecules, etc.) These effects make biochemical computations much harder to design than their electronic counterparts. One way to learn is by characterizing simple biological systems, as was done in a recent comparative study of the "life" of the phage T-7 [10]. The modelers may then be challenged to design different types of artificial phages or even more complicated circuits. Modeling in such contexts is necessarily of type B, since the goal is not to reproduce existing circuits, but to "inspire" new circuits from studying products of evolution. The combination of type-B modeling and gene network synthesis is in our opinion the most fruitful direction to take in the forseeable future, for both the understanding of molecular networks in living cells and the application of these networks to bioengineering (biosensors, environmental detoxification, etc.).

References

[1] P. A. Spiro, J.S. Parkinson, and H.G. Othmer, *PNAS* **94**, 7263-8 (1997).

[2] N. Barkai and S. Leibler, *Nature* **387**, 913-7 (1997).

[3] L. A. Segel, et al., *J. Theor. Biol.* **120**, 151-79 (1986).

[4] D. C. Hauri and J. Ross, *Biophys. J.* **68**, 708-22 (1995).

[5] U. Alon et al., *Nature* **397**, 168-71 (1999).

[6] T.-M. Yi, YHuang, M. I. Simon, and J. Doyle, *PNAS* **97**, 4649-53 (2000).

[7] J. M. Carlson and J. Doyle, *Phys. Rev. E* **60**, 1412-27 (1999).

[8] G. von Dassow, E. Meir, E. M. Munro, and G. M. Odell, *Nature* **406**, 188-92 (2000).

[9] T. S. Gardner, C. R. Cantor, and J. J. Collins, *Nature* **403**, 339-342 (2000).

[10] D. Endy, L. You, J. Yin, I. J. Molineux, Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *PNAS* **10**, 5375-80 (2000).

4.4 Hybrid Neural Network Circuit

As we have indicated, the analysis of genetic circuits is likely to be quite fruitful. There is, of course, a well-developed field devoted to modeling *neural* networks. We note the following example [1] in the analysis of a neural network in biology which illustrates nicely how modeling and simulation can constructively interfere. The system here is actually a neural circuit: the pyloric central pattern generator (CPG) of the lobster whose nervous system is well understood.

The relevant assembly of neurons consists of 14 neurons of which a subassembly consisting of 4 neurons was identified. This subassembly is the critical component of the larger system. The 4 neurons are the anterior burster (AB), two pyloric dilators (PD) and the VD neuron. The methodology was to first measure the activity (intracellular membrane voltage measurements) of the neurons in the assembly, focusing on the neurons of the assembly together with the LP neuron.

Analysis of the data indicted that the neural activity could accurately be described by a dynamical system involving only 3 or 4 dimensions; a detailed 13 dimensional Hodgkin-Huxley model for the neuron was also analyzed to confirm this conclusion. This simplification enabled the actual construction of the electrical neuron (EN).

The AB neuron plays a role in the activity of this neural circuit. Inacti-

vation of this particular neuron destabilized the bursting patterns of the two PD neurons (and so adversely affecting the entire assembly of 14 neurons). Validation of the EN was achieved by electrically "splicing" the EN into the biological circuit (in place of the AB neuron) and demonstrating that in the resulting chimera the bursting patterns of the PD neurons were stabilized, yielding an overall oscillation quite similar to the original pyloric rhythm.

In slightly different situations, neurons have been created "in silico" based on the Hodgkin-Huxley models, however an analog implementation of the AB neuron based on Hodgkin-Huxley models would probably not have been possible with the available resources.

It seems quite plausible that with a library of suitably validated electrical neurons one could explore the possibilities for neural circuits in ways that would not otherwise be possible.

References

[1] A. Szucs, et. al., *Neuroreports* **11**, 563, (2000)

5 RECOMMENDATIONS AND CONCLUSIONS

1. A mix of Type A models, aimed at detailed, quantitative predictions, and Type B models, aimed at abstractions away from details to uncover general principles, is appropriate for modeling of cells. Type B models are more likely to have an impact in the near future, but Type A models have long term prospects for utility in drug testing and as a guide to biological experiments.

2. Modeling efforts should at present concentrate on prokaryotes rather than eukaryotes. Modeling of specific pathways or "modules" is more promising than whole cell modeling even for prokaryotes at the present time.

3. An adequate knowledge of rate constants is a *severe* problem for Type A modeling. The rate constant k_1 in particular (which describes the diffusive processes by which enzymes and substrates meet and dock) is strongly dependent on the cytoplasmic environment. Comparison of the enzyme diffusion constant D_e in vivo vs. in vitro would be a quick test of whether in vitro measurements of k_1 are appropriate to in vivo models of cells. Further studies of the properties of the cytoplasm are clearly needed to address these issues.

4. Models with *spatial* resolution are required for eukaryotes and even for some prokaryotes. The development of such models in turn requires development of experimental methods that can reveal the spatial and temporal biochemical dynamics inside the cell, both to provide input for, and to validate or challenge the models. Such methods might

include CT (computed tomography) scans of cells with soft X-rays, NMR microscopes and standing wave fluoresence microscopy.

5. Dialog and/or collaborations between modelers and the experimental community is critical. Experimentalists are needed to provide meaningful challenges to model builders. Modelers should generate nontrivial falsifiable predictions.

6. Standardized conventions (similar to IUPAC protocols for naming chemical compounds) should be promoted for reporting research on biochemical pathways, especially for representing topology of modules or "circuits".

7. Standardized tests or competitions should be established for "Type A" models. Good models should predict (not "postdict") experimental outcomes, (e.g., results of gene knockouts). The test should be tuned to the organism of interest.

8. "Biochemical circuitry by design" (e.g., toggle switches) via manipulations of gene networks is a particularly interesting area. The time scales are slow ($\sim$ hours) but such experiments are feasible now. Additionally, this kind of technology could be used in principle to bioengineer better "canary" sensors or to program a cell death switch for bacteria after environmental remediation is finished. It is easier to fine tune rate constants in genetic networks than in purely proteomic ones, which is one of the advantages of this approach.

DISTRIBUTION LIST

CMDR & Program Executive Officer
U S Army/CSSD-ZA
Strategic Defense Command
PO Box 15280
Arlington, VA 22215-0150

DARPA
3701 North Fairfax Drive
Arlington, VA 22209-2308

Defense Threat Reduction Agency
Attn: Dr. Arthur T. Hopkins [12]
6801 Telegraph Road
Alexandria,VA 22310

Director of Space and SDI Programs
SAF/AQSC
1060 Air Force Pentagon
Washington, DC 20330-1060

DTIC [2]
8725 John Jay Kingman Road
Suite 0944
Fort Belvoir,VA 22060-6218

Headquarters Air Force XON
4A870
1480 Air Force Pentagon
Washington, DC 20330-1480

Strategic Systems Program
Nebraska Avenue Complex
287 Somers Court
Suite 10041
Washington, DC 20393-5446

Superintendent
Code 1424
Attn Documents Librarian
Naval Postgraduate School
Monterey, CA93943

Dr. Allen Adler
Acting Director
DARPA/TTO
3701 N. Fairfax Drive
Arlington, VA 22209-2308

Dr. Jane Alexander
Acting Director
DARPA/TTO
3701 North Fairfax Drive
Arlington, VA 22203-1714

Dr. A. Michael Andrews
Director of Technology
SARD-TT
Room 3E480
Research Development Acquisition
103 Army Pentagon
Washington, DC 20310-0103

Dr. Albert Brandenstein
Chief Scientist
Office of Nat'l Drug Control Policy
Executive Office of the President
Washington, DC 20500

Dr. H. Lee Buchanan, III
Assistant Secretary of the Navy
(Research, Development & Acquisition)
1000 Navy Pentagon
Washington, DC 20350-1000

Dr. Collier
Chief Scientist
U S Army Strategic Defense Command
PO Box 15280
Arlington, VA 22215-0280

Dr. James F. Decker
Acting Director
Office of Science, SC-1
U.S. Department of Energy
1000 Independence Avenue SW
Washington, DC 20585

Dr. Martin C. Faga
President and Chief Exec Officer
The MITRE Corporation
A210
202 Burlington Road
Bedford,MA 01730-1420

DISTRIBUTION LIST

Mr. Dan Flynn [5]
Program Manager
DI/OT/SAB
DI/OTI/SAG
5 S 49 NHB
Washington, DC 20505

Dr. Paris Genalis
Deputy Director
OUSD(A&T)/S&TS/NW
The Pentagon, Room 3D1048
Washington, DC 20301

Dr. Lawrence K. Gershwin
NIO/S&T
2E42, OHB
Washington, DC 20505

General Thomas F. Gioconda [10]
Acting Deputy Administor
Defense Programs, DP-1
National Nuclear Security Administration
U.S. Department of Energy
1000 Independence Avenue SW
Washington, DC 20585

General John Gordon
Under Secretary for Nuclear Security
U.S. Department of Energy
1000 Independence Avenue SW
Washington, DC 20585

Mr. Lee Hammarstrom
National Reconnaissance Office
14675 Lee Road
Chantilly, VA 20151

Dr. Theodore Hardebeck
STRATCOM/J5B
Offutt AFB, NE68113

Mr. David Havlik
Manager
Weapons Prg Coordination Office
MS 9003
Sandia National Laboratories
PO Box 969
Livermore, CA94551-0969

Dr. Helmut Hellwig
Deputy Asst Secretary
(Science, Technology and Engineering)
SAF/AQR
1060 Air Force Pentagon
Washington, DC 20330-1060

Dr. Robert G. Henderson
Director
JASON Program Office
The MITRE Corporation
1820 Dolley Madison Blvd
McLean,VA 22102

J A S O N Library [5]
The MITRE Corporation
Mail Stop W002
1820 Dolley Madison Blvd
McLeaan,VA 22102

Mr. O' Dean P. Judd
Los Alamos National Laboratory
Mailstop F650
Los Alamos, NM 87545

Dr. Bobby R. Junker
Office of Naval Research
Code 111
800 North Quincy Street
Arlington, VA 22217

Dr. Yeongji Kim
Central Intelligence Agency
Washington, DC 20505

Lt Gen, Howard W. Leaf, (Retired)
Director, Test and Evaluation
HQ USAF/TE
1650 Air Force Pentagon
Washington, DC 20330-1650

Dr. George Mayer
Scientific Director
Army Research Office
4015 Wilson Blvd
Tower 3, Suite 216
Arlington, VA 22203-2529

DISTRIBUTION LIST

Dr. Thomas Meyer
DARPA/DIRO
3701 N. Fairfax Drive
Arlington, VA 22203

Dr. Julian C. Nall
Institute for Defense Analyses
1801 North Beauregard Street
Alexandria, VA 22311

Major General John S. Parker
U.S. Army Medical Research and Material Command
504 Scott Street
Ft Detrick, MD 21702-5012

Dr. Ari Patrinos [5]
Associate Director
Biological and Environmental Research
SC-70
US Department of Energy
19901 Germantown Road
Germantown, MD 207874-1290

Dr. Bruce Pierce
Ballistic Missile Defense Org
BMDO/TO
7100 Defense Pentagon
Room 1E108
Washington, DC 20301-7100

Mr. John Rausch [2]
Division Head 06 Department
NAVOPINCEN
4301 Suitland Road
Washington, DC 20390

Records Resource
The MITRE Corporation
Mail Stop W115
1820 Dolley Madison Blvd
McLean, VA 22102

Dr. Fred E. Saalfeld
Director
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

Dr. Dan Schuresko
Chief
Advanced Technolgoy G roup/
Community Management
Washington, DC 20505

Dr. John Schuster
Submarine Warfare Division
Submarine, Security & Tech
Head (N875)
2000 Navy Pentagon Room 4D534
Washington, DC 20350-2000

Dr. Richard Spinrad
US Naval Observatory
Naval Oceanographers Office
3450 Massachusetts Ave NW
Washington, DC 20392-5421

Dr. Michael A. Stroscio
US Army Research Office
P. O. Box 12211
Research Triangle Park,NC27709-2211

Dr. George W. Ullrich [3]
ODUSD(S&T)
Director for Weapons Systems
3080 Defense Pentagon
Washington, DC 20301-3080

Dr. Linda Zall
Central Intelligence Agency
DS&T/OTS
Washington, DC 20505